



Línea de Trabajo fin de Máster

(Fecha última actualización:...3/3/2022)

Máster en Estadística. CURSO ACADÉMICO 2021-2022	
Título	Desarrollo de un paquete de funciones para generar muestras multivariantes con estructura compleja con propósitos pedagógicos
Tipo	INVESTIGACIÓN <input type="checkbox"/> ORIENTACIÓN PRÁCTICA <input checked="" type="checkbox"/>
Número de alumnos	2
Profesor(es)/ email	Pedro Femia Marzo (pfemia@ugr.es) Antonio Martín Andrés (amartina@ugr.es)
Descripción	<p>Un problema frecuente en la enseñanza de los métodos y técnicas estadísticas relativamente avanzadas, es la necesidad de disponer de muestras de datos que, por una parte, sean adecuadas al contexto pedagógico, y por otra, que reproduzcan de forma satisfactoria aquellas particularidades de la realidad o de la metodología bajo estudio que se desea ilustrar. Por ejemplo, cuando se trata de ajustar modelos que explican la variabilidad de una variable respuesta a través de varias variables explicativas, como es el caso de los modelos de análisis de la varianza o de regresión lineal, es necesario disponer de conjuntos de datos adecuados sobre los que ajustar tales modelos. Adicionalmente, para el estudio metodológico de cada procedimiento estadístico en particular, será necesario contar con conjuntos de datos alternativos. Unos para ser usados como ejemplos, otros como ejercicios propuestos y otros como ejercicios de evaluación.</p> <p>En la práctica, el desarrollo de un curso completo de, por ejemplo, Modelos Lineales, suele requerir de un surtido amplio de bases de datos, cuya recopilación, implementación y adecuación constituye una tarea que demanda mucho tiempo en la preparación del curso. Por otra parte, esos datos recopilados pasan a constituir una fuente de datos estática, en el sentido de que, si bien el conjunto de bases de datos puede ir incrementando con el tiempo, cada matriz siempre estará constituida por los mismos datos.</p> <p>En el plano contextual, en las publicaciones científicas basadas en la observación empírica, es preceptivo encontrar la síntesis estadística de los datos implicados en la investigación junto con los resultados derivados del ajuste de los oportunos modelos explicativos. Sin embargo, lo más frecuente es que los datos originales no sean fácilmente accesibles, y probablemente menos aun cuando lo que se pretende es darles un uso pedagógico. Sin embargo, a partir de la información dada por los autores de una publicación, sí que es posible generar, mediante simulación, unos datos que reproduzcan los resultados dados o bien cualquier otro tipo de aspecto que resulte de interés a la hora de estudiar diferentes situaciones presentes en el fenómeno bajo estudio y la forma de abordarlas analíticamente.</p>
Objetivos particulares	<p>Programación de un paquete de funciones en lenguaje R que permita generar matrices de datos con las relaciones deseadas de dependencia entre las variables que la constituyen. La estructura presente en los datos creados vendrá impuesta por la parametrización indicada en la invocación a la función concreta del paquete, de manera que sea posible reproducir los resultados esperados al analizar estos datos. Es decir, se trata de implementar procedimientos de generación masiva de muestras multivariantes con estructura compleja previamente establecida, apropiadas para poder aplicar sobre ellas técnicas de modelización como pueden ser el análisis de la varianza factorial, el análisis de la covarianza, los modelos con medidas repetidas, modelos de regresión múltiple, regresión logística, etc.</p>



Prerrequisitos y recomendaciones	Conocimientos de inferencia estadística Conocimientos (al menos básicos) de un lenguaje de programación
Plan de trabajo	<ol style="list-style-type: none"> (1) Planteamiento de los esquemas a implementar. En general se abordará la obtención de datos para ser analizados bajo la perspectiva del modelo lineal general. (2) Análisis de requerimientos: esquema de parametrización de los valores de entrada; identificación de las particularidades propias de cada modelo a implementar; análisis de las dependencias respecto a otros paquetes de R. (3) Elaboración del diagrama de flujo y pseudocódigo descriptivo del funcionamiento del paquete (identificación de las funciones a implementar) (4) Programación de las funciones (5) Revisión de errores y puesta a punto del paquete (6) Elaboración de ejemplos: generación de bases de datos e ilustración de su análisis
Competencias generales y específicas	<p>Generales</p> <p>CG1 - Los titulados han de saber aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.</p> <p>CG3 - Los titulados han de saber comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades</p> <p>CG4 - Los titulados deben poseer las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.</p> <p>CG8 - Los titulados deben ser críticos en el análisis, evaluación y síntesis de ideas nuevas y complejas.</p> <p>CG9 - Los titulados deben saber comunicarse con sus colegas, con la comunidad académica en su conjunto y con la sociedad en general acerca de sus áreas de conocimiento.</p> <p>CG10 - Los titulados han de ser capaces de fomentar, en contextos académicos y profesionales, el avance tecnológico, social o cultural dentro de una sociedad basada en el conocimiento.</p> <p>Específicas</p> <p>CE4 - Profundizar en las técnicas de Modelización Estocástica</p> <p>CE9 - Adquirir conocimientos en Estadística</p> <p>CE10 - Dominar el uso de diferentes entornos de Computación Estadística</p>
Bibliografía	<ul style="list-style-type: none"> - Asmussen & Glynn (2007) <i>Stochastic Simulation - Algorithms and Analysis</i>. Springer - Chambers (2008) <i>Software for Data Analysis - Programming with R</i>. Springer - Dobson (2002) <i>An Introduction to Generalized Linear Models</i> (2nd Ed). Chapman & Hall - Dunn & Smyth (2018) <i>Generalized Linear Models with Examples in R</i>. Springer - Voss, J (2014) <i>An Introduction to Statistical Computing. A Simulation-based Approach</i>. Wiley